Preprints.org

Review

Not peer-reviewed version

Generating Test Scenarios for Autonomous Driving: A Taxonomy and Survey

Arif Hossain *

Posted Date: 28 August 2025

doi: 10.20944/preprints202508.2130.v1

Keywords: Simulation; autonomous driving systems; autonomous vehicles

Article

Generating Test Scenarios for Autonomous Driving: A Taxonomy and Survey

Arif Hossain

Independent Researcher; xpdnujzb@pgu.edu.pl

Abstract

Ensuring the safety of autonomous vehicles (AVs) requires rigorous and scalable testing methodologies capable of capturing both routine and safety-critical scenarios. Scenario-based testing has emerged as a vital approach to expose AVs to diverse and challenging conditions beyond traditional road mileage accumulation. This survey focuses on scenario generation—an essential component enabling automated, efficient, and comprehensive testing of autonomous driving systems (ADS). We categorize existing scenario generation methods into three primary paradigms: rule-based, data-driven, and learning-based. For each, we analyze the core methodologies, simulation platforms, scenario description languages, and evaluation metrics used to assess realism, diversity, and criticality. We further identify key research challenges such as the reality gap, limited data generalization, and rare-event modeling, and discuss emerging trends including language-driven generation, hybrid modeling frameworks, and standardized scenario repositories. This work provides a unified perspective on scenario generation, aiming to support researchers and practitioners in advancing safe and certifiable autonomous driving technologies.

Keywords: simulation; autonomous driving systems; autonomous vehicles

---

1. Introduction

Autonomous Driving Systems (ADS) have rapidly evolved in recent years, driven by advances in artificial intelligence, sensor technologies, and vehicular connectivity [1–3]. These systems aim to reduce traffic accidents, enhance transportation efficiency, and improve mobility for various populations. However, ensuring the safety and reliability of ADS before large-scale deployment remains a significant challenge [4–7]. One of the primary barriers is the need for comprehensive testing strategies that can validate system behavior across a wide variety of real-world conditions [8–11].

Testing plays a critical role in the development, validation, and regulatory certification of autonomous vehicles (AVs) [12–15]. Unlike conventional vehicles, AVs operate without direct human intervention and must autonomously perceive and respond to a wide variety of dynamic, uncertain, and high-risk scenarios. These scenarios include interactions with unpredictable human drivers or pedestrians, ambiguous road infrastructure, and rare but safety-critical events, such as sudden occlusions, erratic maneuvers, or sensor degradation [16–18].

Traditional testing approaches that rely heavily on road mileage accumulation-so-called disengagement metrics, are increasingly considered inadequate. Studies have shown that real-world driving may not expose AVs to sufficiently diverse or hazardous edge cases, making it nearly impossible to observe statistically rare failure conditions through brute-force testing alone [19,20]. Moreover, many real-world test kilometers are spent in relatively uneventful scenarios that contribute little to the assessment of system safety [21]. As a result, **scenario-based testing** has gained prominence as a more targeted and efficient alternative. This methodology focuses on designing and executing specific test cases that capture representative driving situations as well as challenging *corner cases* [22,23]. Such scenarios may involve aggressive lane changes, pedestrian dart-outs, occluded intersections, or multi-agent negotiation in dense traffic [24,25].

Scenario-based testing enables both developers and regulators to systematically assess the AV's ability to perceive its environment, make decisions under uncertainty, and maintain safety under predefined conditions. Importantly, these scenarios can be encoded, reused, varied, and scaled through simulation and digital twin environments, ensuring coverage of diverse environmental and behavioral permutations [26–28]. Furthermore, formal scenario generation tools and scenario libraries—such as PEGASUS, ASAM OpenSCENARIO, and Scenic—support the creation of consistent and traceable test cases across different simulation platforms. Regulatory efforts like ISO 21448 (SOTIF) and UNECE WP.29 are also increasingly aligning with scenario-driven safety validation methodologies. Together, these developments mark a shift in AV validation from passive, observational testing to proactive, evidence-driven assessment frameworks.

In this context, **scenario generation** has emerged as a key enabler for efficient, scalable, and safety-oriented testing of autonomous vehicles. Rather than relying solely on manually curated scenarios or recorded driving data, automated scenario generation techniques aim to synthesize diverse, realistic, and safety-critical situations that challenge the capabilities of autonomous driving systems (ADS). These synthetic scenarios are particularly valuable for modeling rare yet high-risk events, such as near-miss collisions, unprotected turns, and aggressive cut-ins,that are statistically unlikely to occur in traditional road testing but essential for evaluating system robustness [29–32].

Recent works have demonstrated the effectiveness of generative approaches, including reinforcement learning, Bayesian optimization [33,34], and Optimization Algorithm [14,35–42], in creating such edge-case scenarios [43–46]. These techniques not only improve testing coverage but also enable *closed-loop simulations* that expose AVs to increasingly difficult decision-making conditions under uncertainty [47–50].

This survey provides a comprehensive review of the state-of-the-art in scenario generation for AV testing. We categorize the methods into three dominant paradigms: **rule-based**, **data-driven**, and **learning-based** scenario generation. For each category, we analyze the algorithmic foundations, simulation tools, scenario description languages (e.g., OpenSCENARIO, Scenic), and evaluation metrics (e.g., coverage, criticality, and diversity). We further identify current limitations, including generalization across environments, reality gaps in simulation, and challenges in modeling multi-agent interactions.

Finally, we explore emerging directions such as *semantic generation via large language models (LLMs)*, interactive multi-agent scenarios, and standardized scenario libraries for benchmarking [51,52]. We believe this survey will serve as a valuable resource for researchers and practitioners striving to build robust, explainable, and certifiable autonomous driving systems.

2. Foundations of Autonomous Driving Testing

Ensuring the safety and functionality of Autonomous Driving Systems (ADS) requires rigorous and systematic testing procedures. Given the complexity and unpredictability of real-world environments, traditional automotive testing methodologies are insufficient to evaluate the full operational domain of AVs. To address this, researchers and regulatory bodies have established a multi-tiered framework for autonomous driving testing, involving various types of environments and structured evaluation methods.

2.1. Testing Categories: Simulation, Closed-Track, and On-Road

Autonomous driving systems (ADS) require rigorous and multi-stage testing pipelines to ensure safety, robustness, and regulatory compliance. Testing strategies are commonly divided into three categories: *simulation-based testing*, *closed-track testing*, and *on-road testing*, each contributing uniquely to the validation landscape.

Simulation-based testing uses virtual environments to replicate driving scenarios under diverse conditions. Tools such as CARLA [53], LGSVL [54], and BeamNG.tech [**?**] enable safe, low-cost evaluation of perception, planning, and control stacks. Simulation allows systematic coverage of rare or extreme scenarios—like sudden occlusions, erratic maneuvers, or unusual lighting—often infeasible in real-world testing [55]. However, the sim-to-real gap in sensor modeling, environment fidelity, and behavioral dynamics remains a key limitation [56,57].

Closed-track testing, conducted in proving grounds or test fields (e.g., Mcity [58], AstaZero [59]), offers controlled and repeatable physical environments. It bridges the realism of on-road testing and the safety of simulation. Critical functionalities such as emergency braking, automated lane changes, and V2X systems can be tested safely [60]. While effective for validating specific behaviors, track testing is expensive and often limited in environmental variety and traffic complexity.

On-road testing is the final and most realistic stage, exposing AVs to real traffic agents, unpredictable events, and regulatory frameworks. Companies like Waymo, Cruise, and Baidu Apollo have reported millions of public testing kilometers. Despite its necessity, on-road testing suffers from poor coverage of rare safety-critical events and carries substantial liability risks [19,61]. Therefore, it is typically preceded by extensive simulation and track validation.

Recent trends advocate for hybrid pipelines that use simulation for exploration, closed-track for validation, and real-world testing for final certification. Table 1 summarizes the trade-offs among the three approaches.

Table 1. Comparison of AV Testing Categories

Category	Advantages	Limitations
Simulation-Based	- Scalable and cost-effective - Safe for humans and vehicles - Enables rare or critical event testing - Fast iteration and automation	- Reality gap in sensor and dynamics fidelity - Synthetic agents oversimplified - Limited realism in weather and edge behaviors
Closed-Track	- Real-world vehicle dynamics - High safety in controlled environments - Repeatable and measurable - Ideal for specific system feature validation	- Infrastructure cost is high - Less diverse scenarios - Not suitable for unstructured environments
On-Road	- Full exposure to real-world traffic - Necessary for regulatory certification - Captures unexpected corner cases	- Safety and liability concerns - Long duration to encounter rare events - Public and legal constraints

2.2. Definitions: Scenario, Scene, and Test Case

In the context of autonomous driving, it is essential to distinguish between several core concepts foundational to testing and scenario generation [62]:

- **Scene**: A *static snapshot* of the driving environment at a specific point in time. It includes information about road layout, infrastructure, traffic participants, weather, lighting conditions, and the dynamic states (e.g., position, velocity, heading) of agents. Scenes represent the spatial and contextual setup without temporal evolution.
- **Scenario**: A *temporal sequence* of scenes that models the unfolding of events involving multiple actors over time. Scenarios define interactions (e.g., overtaking, braking, merging) and enable evaluation of system responses under specific conditions. They are central to behavior modeling and safety testing [63].
- **Test Case**: A *parameterized instantiation* of a scenario, specifying initial configurations (e.g., object positions, speeds, traffic density) and measurable success criteria. Test cases are typically linked with verification outcomes such as pass/fail labels or safety margins.

The hierarchical relationship among scenes, scenarios, and test cases follows the structure proposed in ISO 34502 and is illustrated in Figure 1, where a scenario is composed of a sequence of scenes, and a test case is a specific parameterization of that scenario.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 August 2025 doi:10.20944/preprints202508.2130.v1

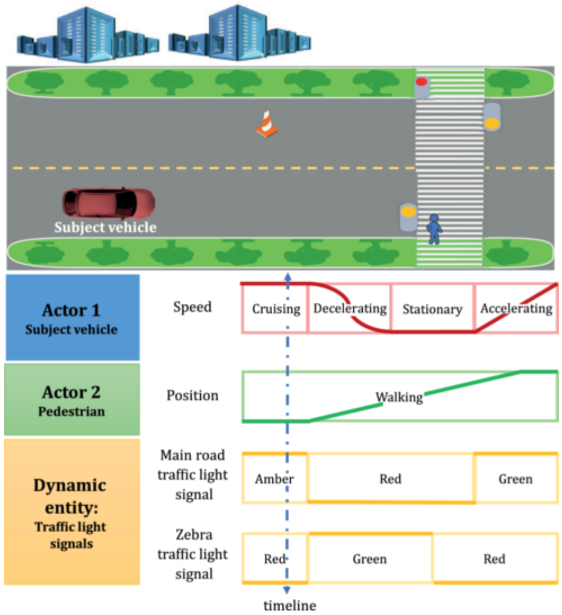

Figure 1. Hierarchical relationship between Scene, Scenario, and Test Case as defined in ISO 34501 [62].

Understanding these distinctions helps formalize testing logic and ensures consistent communication across tools, datasets, and regulatory documentation.

2.3. Common Testing Frameworks and Standards

To enable systematic, reproducible, and certifiable testing of autonomous driving systems (ADS), a number of international standards, industrial frameworks, and research initiatives have emerged. These efforts provide guidelines for scenario modeling, simulation integration, and safety validation processes. We briefly summarize the most widely adopted frameworks and standards below.

PEGASUS Project (Project for the Establishment of Generally Accepted Quality Criteria, Tools and Methods as well as Scenarios and Situations) is a landmark German initiative aiming to formalize scenario-based safety validation of highly automated driving systems. It introduces a structured testing pipeline encompassing scenario derivation from real-world data, parameterization using statistical distributions, simulation-based execution, and coverage-driven evaluation [64]. PEGASUS has become a reference model in Europe for connecting simulation, track, and on-road testing phases with quantitative safety arguments.

ISO 21448, also known as SOTIF (Safety Of The Intended Functionality), addresses safety risks not caused by hardware/software faults (covered by ISO 26262), but by insufficient perception, interpretation, or decision-making in unknown or complex scenarios. It focuses on identifying hazardous behaviors due to limitations of sensing or AI-based reasoning under real-world uncertainties [64**?**]. SOTIF emphasizes hazard identification and mitigation in early-stage functional design and complements failure-based risk management.

ASAM (Association for Standardisation of Automation and Measuring Systems) has developed machine-readable, open standards that support the entire ADS testing workflow:

- **OpenSCENARIO**: Defines dynamic scenario logic, including traffic agents, maneuvers, triggers, and actions.
- **OpenDRIVE**: Describes road network geometry, lanes, signs, and topology for accurate map modeling.
- **OpenXOntology**: Introduces a shared vocabulary for consistent toolchain integration and semantic interoperability.

Other frameworks such as ISO 34501/34502 extend these efforts by defining terminologies, scenario taxonomies, and auditability criteria to ensure traceability and transparency in AV certification.

These frameworks reflect a global shift toward scenario-centric testing pipelines, which enable both proactive hazard exposure and coverage-driven system assurance under realistic, reproducible, and measurable conditions.

2.4. Rule-based Scenario Generation

Rule-based scenario generation methods rely on predefined rules, templates, and domain-specific knowledge to construct test cases. These approaches are often grounded in expert-designed traffic situations, regulatory edge cases, or structured combinations of parameters. They offer high interpretability and are especially suitable for safety-critical scenarios involving well-understood interactions.

A classical approach is to use parameterized templates with domain constraints to produce scenarios covering functional requirements. For example, Klischat and Althoff [65] proposed a method to automatically generate critical scenarios by minimizing the solution space of the ego vehicle's motion planner using evolutionary algorithms such as Differential Evolution and Particle Swarm Optimization. Their framework can effectively generate high-risk scenarios in multi-vehicle interactions and intersections, although computation time increases in high-dimensional spaces due to collision constraints.

Althoff and Lutz [66] further introduced an approach combining reachability analysis with optimization techniques to synthesize collision avoidance scenarios. Their method constructs scenarios within seconds and focuses on ego vehicle safety envelopes but does not yet optimize multi-step trajectories or adversarial agent behavior.

Another line of work addresses scenario diversity and testing efficiency. Feng et al. [67] developed an adaptive testing scenario library generation framework (ATSLG) that leverages Bayesian Optimization with Gaussian Process Regression (GPR) to incrementally refine the scenario space. Their method significantly reduces the number of required test iterations (by 1–2 orders of magnitude) and focuses on critical scenario discovery, though its performance in high-dimensional feature spaces still requires enhancement.

Gao et al. [68] proposed a combinatorial test generation strategy combining test matrices and a complexity-driven combination algorithm (CTBC). Their method incorporates AHP-based hierarchical influence modeling to balance scenario coverage and complexity. While the approach improves scenario quality and test defect detection under budget constraints, its complexity estimation relies on approximations and bounded assumptions.

Zhang et al. [69] introduced a new direction by incorporating knowledge-enhanced scenario synthesis via LLMs, aligning natural language intent with parameterized generation, though the method builds upon and extends traditional rule-based backbones.

Overall, rule-based methods are widely adopted due to their control, repeatability, and traceability, making them attractive for regulatory testing and safety assurance. However, they face challenges in capturing the long-tail of rare or emergent interactions and often lack adaptability in open-ended environments.

2.5. Data-driven Scenario Generation

Data-driven scenario generation leverages large-scale naturalistic driving datasets to extract, learn, and synthesize new scenarios that capture realistic traffic patterns, behaviors, and edge cases. This approach bypasses manual modeling by utilizing statistical, learning-based, or heuristic techniques to infer underlying dynamics from real-world data. The richness and diversity of existing datasets such as NGSIM, Argoverse, or SHRP2 enable scalable creation of test cases that mirror complex real-world situations.

Trajectory-based learning has emerged as a common strategy in this category. Zhang et al. [70] proposed *DP-TrajGAN*, a generative adversarial framework augmented with differential privacy to synthesize high-fidelity trajectories. Their method balances utility and privacy, and was validated on NGSIM and Argoverse datasets. Similarly, Krajewski et al. [71] combined GANs and VAEs to model realistic vehicle maneuvers, supporting scenario diversity and simulation accuracy.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 August 2025 doi:10.20944/preprints202508.2130.v1

To increase the criticality of generated scenarios, adversarial perturbation techniques have also been explored. Wang et al. [72] introduced *AdvSim*, which modifies vehicle trajectories and LiDAR signals to create safety-critical conditions. Their framework demonstrated the capability to uncover system weaknesses through adversarial replay in simulation.

Beyond trajectory generation, several studies have focused on **quantifying and targeting criticality**. Westhofen et al. [73] conducted a comprehensive review of criticality metrics and proposed a framework for assessing their suitability in various testing contexts. Kang et al. [74] employed voxel-based 3D modeling and vision transformers to detect latent safety threats in LiDAR data, achieving a high F1 score (98.26%) in identifying risky zones.

Driving instability as a data-driven indicator of crash likelihood has also been studied. Arvin et al. [75] analyzed SHRP2 data and confirmed the correlation between pre-crash instability and crash severity, which can guide the generation of high-risk situations.

Rare and corner-case detection is another active thread. Bolte et al. [76] proposed a framework that detects low-frequency, high-impact events by combining offline anomaly detection with online event flagging, enabling automatic harvesting of rare scenario candidates from driving logs.

Parameter-based scenario abstraction was proposed by Muslim et al. [77], who generated cut-out scenarios from highway data using interpretable parameter boundaries. This method ensures scenario plausibility while maintaining control over scenario variation.

From a controllability and diversity perspective, Huang et al. [78] introduced the *CaDRE* framework, which fuses real-world trajectory distributions with quality-diversity optimization to generate representative and safety-critical scenarios.

2.6. Learning-based Scenario Generation

Learning-based scenario generation represents the most recent and dynamic frontier. These methods leverage machine learning, especially generative models, to synthesize novel and complex scenarios beyond the scope of curated data.

Generative Adversarial Networks (GANs) and **Variational Autoencoders (VAEs)** have been applied to generate traffic scenarios with controlled agent behavior, spatial configurations, and environmental features. Recently, **diffusion models** have shown promise in generating temporally coherent trajectories and multi-agent interactions.

Another active direction is **reinforcement learning (RL)**, where an adversarial agent is trained to generate scenarios that maximize the failure likelihood of the system under test. These *failure-triggering* scenarios are valuable for testing system robustness under stress conditions. Some methods use closed-loop feedback, where scenario generation is iteratively optimized based on AV performance metrics, creating adaptive testing pipelines.

Learning-based approaches offer strong potential for scalability, diversity, and automation. However, they also introduce challenges such as scenario validity, safety assurance, and interpretability.

3. Scenario Quality and Evaluation Metrics

A critical aspect of scenario-based testing for autonomous driving lies not only in generating test scenarios but also in evaluating their quality and relevance. To ensure that test scenarios contribute meaningfully to system validation, researchers have proposed several quality criteria and quantitative metrics. These metrics assess how well the scenarios represent real-world driving conditions, how diverse and critical they are, and how effectively they expose weaknesses in the system under test.

Evaluation Criteria

Scenario evaluation typically revolves around five core dimensions:

- **Realism**: Measures how closely a generated scenario resembles those observed in real-world driving. This includes plausible agent behavior, realistic motion dynamics, and compliance with

traffic rules. Realism ensures external validity and is often evaluated using human annotation or statistical comparison to naturalistic datasets [79–81].
- **Diversity**: Refers to the breadth of variation among generated scenarios, covering different maneuvers, traffic densities, environmental conditions, and agent interactions. High diversity increases the likelihood of discovering unforeseen failure modes [82–85].
- **Coverage**: Describes how well the generated scenarios span the space of operational design domain (ODD) conditions and functional safety requirements. Coverage can be quantified via semantic tags (e.g., highway merging, unprotected left turns) or parameter space sampling metrics [37,40,84,86].
- **Criticality**: Indicates the level of challenge or risk presented by a scenario. Metrics for criticality include time-to-collision (TTC), minimum distance, deceleration demand, or probability of collision. These help prioritize scenarios likely to reveal unsafe behavior [87–91].

4. Simulation Platforms and Scenario Description Languages

Effective evaluation of autonomous driving systems (ADS) requires realistic, flexible, and extensible simulation environments. These simulators enable controlled testing of perception, planning, and control components in a safe and scalable manner. Complementing these platforms, scenario description languages provide structured mechanisms to encode and manipulate the test conditions and agent behaviors. Together, simulators and scenario languages form the backbone of scenario-based testing pipelines.

4.1. Key Simulation Environments

Several high-fidelity simulation platforms have been developed to support AV research, each with different emphases on realism, customizability, and system integration.

- **CARLA (Car Learning to Act)**: An open-source simulator designed for AV research, CARLA supports sensor simulation (RGB, LiDAR, radar), weather variation, and custom traffic scenarios. It integrates well with reinforcement learning agents and scenario definitions via Python APIs. CARLA supports OpenDRIVE for map import and OpenSCENARIO (experimental) for scenario control.
- **LGSVL (now SVL Simulator)**: Built on Unity3D, LGSVL offers photorealistic environments and detailed physics, supporting multiple AV stacks such as Apollo and Autoware. It provides APIs for ego-vehicle control and external scenario integration.
- **Apollo Simulation Platform**: As part of Baidu's Apollo open-source AV stack, this platform provides integration-ready tools for sensor simulation, cyber modules, and scenario playback. It is particularly well-suited for closed-loop system validation.

These platforms enable scalable and repeatable testing while supporting a wide range of use cases, from low-level control testing to high-level decision-making evaluation.

4.2. Scenario Languages and Tools

To specify and control complex test scenarios across simulators, several scenario description languages and tools have been developed:

- **OpenSCENARIO**: A widely adopted standard developed by ASAM, OpenSCENARIO defines scenarios in an XML format including actors, events, triggers, and environment settings. It supports both deterministic and stochastic scenario execution and is interoperable with tools such as OpenDRIVE and VTD.
- **Scenic**: A probabilistic programming language for scenario specification, Scenic allows concise descriptions of scenes using constraints and distributions. It is well-suited for generating diverse and controlled scenes for simulation, especially in platforms like CARLA.

- **SceneDSL**: A domain-specific language developed for modular and reusable scenario design. It allows hierarchical definitions of behaviors, goals, and events, facilitating large-scale scenario generation with abstraction and code reuse.

Each language offers different levels of expressiveness, modularity, and support for randomness, and the choice often depends on the complexity and variability requirements of the test scenarios.

5. Challenges and Research Gaps

Despite significant progress in scenario generation for autonomous driving testing, several challenges and open research questions remain. These issues span technical, practical, and regulatory domains, highlighting the need for more robust, scalable, and standardized solutions. This section outlines the most critical challenges faced by current approaches and identifies gaps that present opportunities for future work.

Limited Data Diversity and Generalization

Most data-driven and learning-based generation techniques rely on real-world driving datasets, which, while extensive, often exhibit distributional bias. They tend to overrepresent common urban driving patterns and underrepresent edge cases such as near-misses, rare weather conditions, or unusual traffic interactions. As a result, generated scenarios may lack diversity and fail to generalize across unseen domains or geographies. Bridging this gap requires improved domain adaptation methods, cross-city or cross-country datasets, and techniques for transferring knowledge between different driving contexts.

Reality Gap in Synthetic Scenarios

A persistent challenge is the so-called *reality gap*—the mismatch between synthetic, simulator-generated scenarios and real-world driving environments. Even high-fidelity simulators may fail to capture subtle behaviors of pedestrians, occlusions, sensor noise, or infrastructure imperfections. This gap can lead to overestimation of AV performance in simulated testing and underpreparedness in real-world deployment. Addressing this issue involves combining real-world and simulated data, applying domain randomization, and improving simulator realism both at the perception and decision-making levels.

Scalability of Scenario Space

The scenario space for AV testing is practically infinite, encompassing a large number of interacting variables—agent types, behaviors, road layouts, environmental conditions, and temporal sequences. Exhaustive exploration of this space is infeasible. Thus, existing generation methods often sample from constrained subspaces, risking incomplete validation. New scalable methods for scenario space abstraction, semantic scenario clustering, and combinatorial coverage optimization are needed to ensure high testing efficiency without sacrificing thoroughness.

Modeling Safety-Critical but Rare Events

Safety-critical scenarios—such as sudden pedestrian crossings, aggressive merges, or multi-agent collisions—are rare but essential for validating robustness. However, they are underrepresented in data and difficult to model without manual engineering or adversarial optimization. Existing methods either rely on hand-crafted triggers or heuristic optimization, both of which may fail to cover unknown failure modes. Learning methods that can autonomously discover and amplify rare, high-risk patterns are still in their infancy and represent a vital area for advancement.

Standardization and Regulatory Alignment

A major obstacle in the deployment of scenario-based testing is the lack of unified standards for scenario modeling, evaluation, and exchange. While efforts like ASAM OpenSCENARIO and ISO 21448 (SOTIF) provide foundational standards, their integration into automated generation pipelines

doi:10.20944/preprints202508.2130.v1

remains limited. Furthermore, regulatory acceptance of synthetic and generated scenarios for AV certification is still evolving. Research is needed on formal scenario verification, traceability, and compliance to ensure generated scenarios meet safety assurance requirements in a legally verifiable way.

6. Emerging Trends and Future Directions

As autonomous driving technologies mature, the demands on scenario generation frameworks are becoming more sophisticated. To support next-generation AV testing, recent research efforts have begun to explore novel paradigms that go beyond conventional rule-based or data-centric approaches. This section highlights several emerging trends and outlines promising future directions that are expected to shape the field in the coming years.

Semantic and Language-Driven Scenario Generation

The integration of large language models (LLMs) such as GPT and Claude into the AV testing pipeline has opened a new frontier: **semantic scenario generation**. Instead of specifying low-level scene parameters manually or learning them from data, users can now describe high-level scenarios in natural language—e.g., “a pedestrian suddenly crosses at a dimly lit intersection during rain”—which are then automatically translated into structured scenario code (e.g., OpenSCENARIO or Scenic). This paradigm enables more intuitive, human-centered interaction and lowers the barrier for specifying complex or rare situations.

Several early-stage systems now link language models with simulation backends (e.g., CARLA + LLM), enabling real-time scenario synthesis and editing. Future work may focus on integrating commonsense reasoning, legal constraints, and safety specifications directly into the generation pipeline through natural language interfaces.

Multi-modal and Multi-agent Scene Synthesis

Another active trend is the development of **multi-modal scenario synthesis** that incorporates visual, spatial, and behavioral information from multiple sources—such as video, LiDAR, maps, and text—to construct comprehensive test scenes. Generative models are being trained to combine these modalities into coherent environments, which better reflect the sensor fusion-based perception systems in real AVs.

In parallel, there is increasing interest in **multi-agent interaction modeling**. Modern urban scenarios often involve complex interactions among multiple agents with varying intent (e.g., pedestrians, cyclists, autonomous and human-driven vehicles). Modeling these interactions realistically, and generating coordinated behavior trajectories, remains a significant challenge. Multi-agent reinforcement learning, game-theoretic approaches, and diffusion-based generative models are emerging tools for tackling this complexity.

Hybrid Data-Driven and Rule-Based Approaches

Recognizing the limitations of pure rule-based or data-driven methods, researchers are moving towards **hybrid frameworks** that combine both strengths. Rule-based constraints provide safety and structure, while data-driven models contribute realism and diversity.

In practice, this might involve using data-driven models to sample base scenes and agents, with rule-based logic applied to inject specific intent, constraints, or triggers. Alternatively, hybrid approaches may operate in a layered architecture—where a symbolic planner outlines scenario semantics, and a learned module fills in the low-level details. These combinations are particularly promising for balancing interpretability with expressive power.

Towards Standardized Scenario Repositories and Benchmarks

As the field grows, there is a growing need for **open, standardized scenario repositories and benchmarking protocols**. Currently, many datasets and scenarios are either proprietary or fragmented, making reproducibility and comparative evaluation difficult.

Initiatives such as the **ASAM OpenX** family (OpenSCENARIO, OpenDRIVE, OpenLABEL) and projects like **GENEVA** or **SAFETAG** aim to unify scenario description formats and provide comprehensive libraries of validated test cases. Benchmarking tools that evaluate scenario quality, failure exposure, and coverage are also under active development.

In the future, publicly maintained scenario banks—similar to ImageNet for vision or GLUE for NLP—may become the cornerstone for training, testing, and certifying autonomous driving systems under global safety standards.

7. Conclusion

Scenario generation has emerged as a central pillar in the validation and verification of autonomous driving systems. This survey has reviewed the landscape of scenario generation techniques from multiple perspectives, including rule-based, data-driven, and learning-based approaches. We have discussed key simulation platforms and scenario description languages that underpin automated testing pipelines, as well as the metrics used to evaluate scenario quality in terms of realism, diversity, criticality, and reproducibility.

From the reviewed literature and practices, several important takeaways can be drawn. Rule-based generation provides structure and standardization but struggles with diversity and scalability. Data-driven approaches benefit from realism grounded in real-world observations but are constrained by dataset limitations and rarity of critical events. Learning-based methods offer promising adaptability and automation, especially for generating adversarial or failure-triggering scenarios, but face challenges related to safety, interpretability, and validation.

Scenario quality evaluation remains a non-trivial task, requiring multidimensional metrics and feedback from simulation environments. Tools like Scenic and general coverage metrics are gaining traction in quantifying scenario space exploration. Moreover, the growing integration of scenario generation with closed-loop simulators enables dynamic, intelligent testing strategies that evolve alongside AV system development.

Looking forward, advancing scenario generation will be key to achieving safe and efficient autonomous driving. Future research should prioritize hybrid and semantic methods that balance structure with adaptability, develop standardized scenario libraries and benchmarking protocols, and close the realism gap between synthetic and real-world driving environments. Through collaborative efforts in research, tool development, and regulation, scenario-based testing will continue to evolve as a robust framework for ensuring safety in increasingly complex autonomous systems.

References

1. Zhang, Q.; Hua, K.; Zhang, Z.; Zhao, Y.; Chen, P. ACNet: An Attention–Convolution Collaborative Semantic Segmentation Network on Sensor-Derived Datasets for Autonomous Driving. *Sensors* **2025**, *25*, 4776.
2. Wei, Z.; Gutierrez, C.A.; Rodr, J.; Wang, J. 6G-enabled Vehicle-to-Everything Communications: Current Research Trends and Open Challenges. *IEEE Open Journal of Vehicular Technology* **2025**.
3. Kumar, H.; Mamoria, P.; Dewangan, D.K. Vision technologies in autonomous vehicles: progress, methodologies, and key challenges. *International Journal of System Assurance Engineering and Management* **2025**.
4. Liu, X.; Huang, H.; Bian, J.; Zhou, R.; Wei, Z. Generating intersection pre-crash trajectories for autonomous driving safety testing using Transformer Time-Series GANs. *Engineering Applications of Artificial Intelligence* **2025**.
5. Zhao, M.; Liang, C.; Wang, T.; Guan, J.; Wan, L. Scenario Hazard Prevention for Autonomous Driving Based on Improved STPA. In *Safety, Reliability, and Security*; Springer, 2025.

6. Zhou, R.; Huang, H.; Lee, J.; Huang, X.; Chen, J.; Zhou, H. Identifying typical pre-crash scenarios based on in-depth crash data with deep embedded clustering for autonomous vehicle safety testing. *Accident Analysis & Prevention* **2023**, *191*, 107218.
7. Huang, H.; Huang, X.; Zhou, R.; Zhou, H.; Lee, J.J.; Cen, X. Pre-crash scenarios for safety testing of autonomous vehicles: A clustering method for in-depth crash data. *Accident Analysis & Prevention* **2024**, *203*, 107616.
8. da Costa, A.A.B.; Irvine, P.; Dodoiu, T.; Khastgir, S. Building a Robust Scenario Library for Safety Assurance of Automated Driving Systems: A Review. *IEEE Transactions on Intelligent Transportation Systems* **2025**.
9. Lahikainen, J. AI-Driven Inverse Method for Identifying Mechanical Properties From Small Punch Tests. PhD thesis, Aalto University, 2025.
10. Zhou, R.; Zhang, G.; Huang, H.; Wei, Z.; Zhou, H.; Jin, J.; Chang, F.; Chen, J. How would autonomous vehicles behave in real-world crash scenarios? *Accident Analysis & Prevention* **2024**, *202*, 107572.
11. Gu, J.; Bellone, M.; Lind, A. Camera-LiDAR Fusion based Object Segmentation in Adverse Weather Conditions for Autonomous Driving. In Proceedings of the 2024 19th Biennial Baltic Electronics Conference (BEC). IEEE, 2024.
12. Feng, S.; Sun, H.; Yan, X.; Zhu, H.; Zou, Z.; Shen, S.; Liu, H.X. Dense reinforcement learning for safety validation of autonomous vehicles. *Nature* **2023**, *615*, 620–627.
13. Feng, S.; Yan, X.; Sun, H.; Feng, Y.; Liu, H.X. Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment. *Nature communications* **2021**, *12*, 748.
14. Huang, X.; Cen, X.; Cai, M.; Zhou, R. A framework to analyze function domains of autonomous transportation systems based on text analysis. *Mathematics* **2022**, *11*, 158.
15. Scanlon, J.M.; Kusano, K.D.; Daniel, T.; Alderson, C.; Ogle, A.; Victor, T. Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain. *Accident Analysis & Prevention* **2021**, *163*, 106454.
16. Huang, W.; Wang, K.; Lv, Y.; Zhu, F. Autonomous vehicles testing methods review. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2016, pp. 163–168.
17. Hungar, H. Scenario-based validation of automated driving systems. In Proceedings of the International Symposium on Leveraging Applications of Formal Methods. Springer, 2018, pp. 449–460.
18. Deldari, N. Scenario Annotation in Autonomous Driving: An Outlier Detection Framework. PhD thesis, Uppsala University, 2025.
19. Kalra, N.; Paddock, S.M. Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability? *Transportation Research Part A* **2016**.
20. Ding, W.; Lin, H.; Li, B.; Zhao, D. Causalaf: Causal autoregressive flow for safety-critical driving scenario generation. In Proceedings of the Conference on robot learning. PMLR, 2023, pp. 812–823.
21. Li, C.; Sifakis, J.; Wang, Q.; Yan, R.; Zhang, J. Simulation-based validation for autonomous driving systems. In Proceedings of the Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis, 2023, pp. 842–853.
22. Nalic, D.; Mihalj, T.; Bäumler, M.; Lehmann, M.; Eichberger, A.; Bernsteiner, S. Scenario based testing of automated driving systems: A literature survey. In Proceedings of the FISITA web Congress, 2020, Vol. 10.
23. Sahu, N.; Bhat, A.; Rajkumar, R. SafeRoute: Risk-Minimizing Cooperative Real-Time Route and Behavioral Planning for Autonomous Vehicles. In Proceedings of the 2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2024.
24. Nagy, R.; Szalai, I. Development of an unsupervised learning-based annotation method for road quality assessment. *Transportation Engineering* **2025**.
25. Menzel, T.; Bagschik, G.; Maurer, M. Scenarios for development, test and validation of automated vehicles. In Proceedings of the 2018 IEEE intelligent vehicles symposium (IV). IEEE, 2018, pp. 1821–1827.
26. Menzel, T.; Bagschik, M.; Maurer, M. Scenarios for Safety Validation of Highly Automated Vehicles. *Transportation Research Part F* **2020**.
27. Sun, J.; Zhang, H.; Zhou, H.; Yu, R.; Tian, Y. Scenario-based test automation for highly automated vehicles: A review and paving the way for systematic safety assurance. *IEEE transactions on intelligent transportation systems* **2021**, *23*, 14088–14103.
28. Fiorino, M.; Naeem, M.; Ciampi, M.; Coronato, A. Defining a metric-driven approach for learning hazardous situations. *Technologies* **2024**, *12*, 103.

 doi:10.20944/preprints202508.2130.v1

29. Mammen, M.; Kayatas, Z.; Bestle, D. Evaluation of Different Generative Models to Support the Validation of Advanced Driver Assistance Systems. *Applied Mechanics* **2025**.
30. Zhang, X.; Xiong, L.; Zhang, P.; Huang, J.; Ma, Y. Real-world Troublemaker: A 5G Cloud-controlled Track Testing Framework for Automated Driving Systems in Safety-critical Interaction Scenarios. *arXiv preprint arXiv:2502.14574* **2025**.
31. Wang, J.; Wang, X. Safety-Critical Scenario Generation for Self-Driving Systems Based on Domain Models. In Proceedings of the 2nd International Conference on Intelligent Robotics and Control Engineering, 2025.
32. Zhou, R.; Huang, H.; Lee, J.; Huang, X.; Chen, J.; Zhou, H. Identifying typical pre-crash scenarios based on in-depth crash data with deep embedded clustering for autonomous vehicle safety testing. *Accident Analysis & Prevention* **2023**, *191*, 107218.
33. Xu, C.; Ding, Z.; Wang, C.; Li, Z. Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes. *Journal of Safety Research* **2019**, *71*, 41–47.
34. Lenard, J. Typical pedestrian accident scenarios for the development of autonomous emergency braking test protocols. *Accident Analysis & Prevention* **2014**, p. 8.
35. Cai, J.; Zhou, R.; Lei, D. Dynamic shuffled frog-leaping algorithm for distributed hybrid flow shop scheduling with multiprocessor tasks. *Engineering Applications of Artificial Intelligence* **2020**, *90*, 103540.
36. Zhou, R.; Liu, Y.; Zhang, K.; Yang, O. Genetic algorithm-based challenging scenarios generation for autonomous vehicle testing. *IEEE Journal of Radio Frequency Identification* **2022**, *6*, 928–933.
37. Zhu, B.; Sun, Y.; Zhao, J.; Han, J.; Zhang, P.; Fan, T. A critical scenario search method for intelligent vehicle testing based on the social cognitive optimization algorithm. *IEEE Transactions on Intelligent Transportation Systems* **2023**, *24*, 7974–7986.
38. Zhou, R. Efficient Safety Testing of Autonomous Vehicles via Adaptive Search over Crash-Derived Scenarios. *arXiv preprint arXiv:2508.06575* **2025**.
39. Zhou, R.; Lei, D.; Zhou, X. Multi-objective energy-efficient interval scheduling in hybrid flow shop using imperialist competitive algorithm. *IEEE access* **2019**, *7*, 85029–85041.
40. Bian, J.; Huang, H.; Yu, Q.; Zhou, R. Search-to-Crash: Generating safety-critical scenarios from in-depth crash data for testing autonomous vehicles. *Energy* **2025**, p. 137174.
41. Sun, J.; Zhang, H.; Zhou, H.; Yu, R.; Tian, Y. Scenario-Based Test Automation for Highly Automated Vehicles: A Review and Paving the Way for Systematic Safety Assurance. *IEEE Transactions on Intelligent Transportation Systems* **2022**, *23*, 14088–14103. https://doi.org/10.1109/TITS.2021.3136353.
42. Tian, Y.; Zheng, W.; Shao, Y.; Zhang, H.; Sun, J. MJTG: A Multi-vehicle Joint Trajectory Generator for Complex and Rare Scenarios. *IEEE Transactions on Vehicular Technology* **2025**.
43. Mondelli, A.; Li, Y.; Zanardi, A.; Frazzoli, E. Test Automation for Interactive Scenarios via Promptable Traffic Simulation. *arXiv preprint arXiv:2506.01199* **2025**.
44. Mei, Y.; Nie, T.; Sun, J.; Tian, Y. Llm-attacker: Enhancing closed-loop adversarial scenario generation for autonomous driving with large language models. *arXiv preprint arXiv:2501.15850* **2025**.
45. Zeng, Z.; Shi, Q.; Zhuang, W.; Wang, X.; Fan, X. Adversarial Generation for Autonomous Vehicles in Safety-Critical Ramp Merging Scenarios. In Proceedings of the International Conference on Electric Vehicle and Vehicle Engineering. Springer, 2024, pp. 427–434.
46. Feng, S.; Feng, Y.; Sun, H.; Zhang, Y.; Liu, H.X. Testing scenario library generation for connected and automated vehicles: An adaptive framework. *IEEE Transactions on Intelligent Transportation Systems* **2020**, *23*, 1213–1222.
47. Tang, S.; Zhang, Z.; Zhou, J.; Lei, L.; Zhou, Y.; Xue, Y. Legend: A top-down approach to scenario generation of autonomous driving systems assisted by large language models. In Proceedings of the Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering, 2024, pp. 1497–1508.
48. Arnav, M. Scenario generation methods for functional safety testing of automated driving systems **2025**.
49. Huai, Y.; Almanee, S.; Chen, Y.; Wu, X.; Chen, Q.A.; Garcia, J. scenoRITA: Generating Diverse, Fully-Mutable, Test Scenarios for Autonomous Vehicle Planning. *IEEE Transactions on Software Engineering* **2023**, pp. 1–21.
50. Ding, W.; Lin, H.; Li, B.; Zhao, D. Generalizing goal-conditioned reinforcement learning with variational causal reasoning. *Advances in Neural Information Processing Systems* **2022**, *35*, 26532–26548.
51. Cai, X.; Bai, X.; Cui, Z.; Xie, D.; Fu, D.; Yu, H.; Ren, Y. Text2scenario: Text-driven scenario generation for autonomous driving test. *arXiv preprint arXiv:2503.02911* **2025**.
52. Ricotta, C.; Khzym, S.; Faron, A.; Emadi, A. Property Optimized GNN: Improving Data Association Performance Using Cost Function Optimization for Sensor Fusion In High Density Environments. In Proceedings of the 2024 IEEE Smart World Congress (SWC). IEEE, 2024, pp. 1871–1877.

 doi:10.20944/preprints202508.2130.v1

53. Dosovitskiy, A.; Ros, G.; Codevilla, F.; Lopez, A.; Koltun, V. CARLA: An open urban driving simulator. In Proceedings of the Conference on robot learning. PMLR, 2017, pp. 1–16.
54. Rong, G.; Shin, B.H.; Tabatabaee, H.; Lu, Q.; Lemke, S.; Možeiko, M.; Boise, E.; Uhm, G.; Gerow, M.; Mehta, S.; et al. Lgsvl simulator: A high fidelity simulator for autonomous driving. In Proceedings of the 2020 IEEE 23rd International conference on intelligent transportation systems (ITSC). IEEE, 2020, pp. 1–6.
55. Maier, R.; Grabinger, L.; Urlhart, D.; Mottok, J. Causal models to support scenario-based testing of adas. *IEEE Transactions on Intelligent Transportation Systems* **2023**, *25*, 1815–1831.
56. Fremont, D.J.; Kim, E.; Pant, Y.V.; Seshia, S.A.; Acharya, A.; Bruso, X.; Wells, P.; Lemke, S.; Lu, Q.; Mehta, S. Formal scenario-based testing of autonomous vehicles: From simulation to the real world. In Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2020, pp. 1–8.
57. Sonetta, H.Y. Bridging the simulation-to-reality gap: Adapting simulation environment for object recognition. Master's thesis, University of Windsor (Canada), 2021.
58. Dong, Y.; Zhong, Y.; Yu, W.; Zhu, M.; Lu, P.; Fang, Y.; Hong, J.; Peng, H. Mcity data collection for automated vehicles study. *arXiv preprint arXiv:1912.06258* **2019**.
59. Jacobson, J.; Janevik, P.; Wallin, P. Challenges in creating AstaZero, the active safety test area. In Proceedings of the Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to DeploymentEuropean CommissionConference of European Directors of Roads (CEDR) European Road Transport Research Advisory Council (ERTRAC), 2014.
60. Ma, Y.; Sun, C.; Chen, J.; Cao, D.; Xiong, L. Verification and validation methods for decision-making and planning of automated vehicles: A review. *IEEE Transactions on Intelligent Vehicles* **2022**, *7*, 480–498.
61. Mariani, R. An overview of autonomous vehicles safety. In Proceedings of the 2018 IEEE International Reliability Physics Symposium (IRPS). IEEE, 2018, pp. 6A–1.
62. International Organization for Standardization. ISO 34501:2022 - Road vehicles — Test scenarios for automated driving systems — Vocabulary, 2022.
63. Batsch, F.; Kanarachos, S.; Cheah, M.; Ponticelli, R.; Blundell, M. A taxonomy of validation strategies to ensure the safe operation of highly automated vehicles. *Journal of Intelligent Transportation Systems* **2021**, *26*, 14–33.
64. Wang, C.; Storms, K.; Zhang, N.; Winner, H. Runtime unknown unsafe scenarios identification for SOTIF of autonomous vehicles. *Accident Analysis & Prevention* **2024**, *195*, 107410.
65. Klischat, M.; Althoff, M. Generating critical test scenarios for automated vehicles with evolutionary algorithms. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), 2019.
66. Althoff, M.; Lutz, S. Automatic generation of safety-critical test scenarios for collision avoidance of road vehicles. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), 2018.
67. Feng, S.; Feng, Y.; Sun, H.; Zhang, Y.; Liu, H.X. Testing scenario library generation for connected and automated vehicles: An adaptive framework. *IEEE Transactions on Intelligent Transportation Systems* **2020**.
68. Gao, F.; Duan, J.; He, Y.; Wang, Z. A test scenario automatic generation strategy for intelligent driving systems. *Mathematical Problems in Engineering* **2019**.
69. Zhang, J.; Xu, C.; Li, B. Chatscene: Knowledge-enabled safety-critical scenario generation for autonomous vehicles. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024.
70. Zhang, J.; Huang, Q.; Huang, Y. DP-TrajGAN: A privacy-aware trajectory generation model with differential privacy. *Future Generation Computer Systems* **2023**.
71. Krajewski, R.; Moers, T.; Nerger, D. Data-driven maneuver modeling using generative adversarial networks and variational autoencoders for safety validation of highly automated vehicles. In Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC), 2018.
72. Wang, J.; Pun, A.; Tu, J. Advsim: Generating safety-critical scenarios for self-driving vehicles. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
73. Westhofen, L.; Neurohr, C.; Koopmann, T. Criticality metrics for automated driving: A review and suitability analysis of the state of the art. *Archives of Computational Methods in Engineering* **2023**.
74. Kang, M.; Seo, J.; Hwang, K. Critical voxel learning with vision transformer and derivation of logical AV safety assessment scenarios. *Accident Analysis & Prevention* **2024**.
75. Arvin, R.; Kamrani, M.; Khattak, A.J. The role of pre-crash driving instability in contributing to crash intensity using naturalistic driving data. *Accident Analysis & Prevention* **2019**.
76. Bolte, J.A.; Bar, A.; Lipinski, D. Towards corner case detection for autonomous driving. In Proceedings of the IEEE Intelligent Vehicles Symposium, 2019.

77. Muslim, H.; Endo, S.; Imanaga, H. Cut-out scenario generation with reasonability foreseeable parameter range from real highway dataset for autonomous vehicle assessment. *IEEE Access* **2023**.
78. Huang, P.; Ding, W.; Francis, J. CaDRE: Controllable and Diverse Generation of Safety-Critical Driving Scenarios using Real-World Trajectories. *arXiv preprint arXiv:2401.XXXX* **2024**.
79. Zhou, R.; Lin, Z.; Zhang, G.; Huang, H.; Zhou, H.; Chen, J. Evaluating autonomous vehicle safety performance through analysis of pre-crash trajectories of powered two-wheelers. *IEEE Transactions on Intelligent Transportation Systems* **2024**, *25*, 13560–13572.
80. Zhou, R.; Gui, W.; Huang, H.; Liu, X.; Wei, Z.; Bian, J. DiffCrash: Leveraging Denoising Diffusion Probabilistic Models to Expand High-Risk Testing Scenarios Using In-Depth Crash Data. *Expert Systems with Applications* **2025**, p. 128140.
81. Zhang, G.; Huang, H.; Zhou, R.; Li, S.; Bian, J. High-Risk Trajectories Generation for Safety Testing of Autonomous Vehicles Based on In-Depth Crash Data. *IEEE Transactions on Intelligent Transportation Systems* **2025**.
82. Oliveira, B.B.; Carravilla, M.A.; Oliveira, J.F. A diversity-based genetic algorithm for scenario generation. *European journal of operational research* **2022**, *299*, 1128–1141.
83. Chu, Q.; Yue, Y.; Yao, D.; Pei, H. DiCriTest: Testing Scenario Generation for Decision-Making Agents Considering Diversity and Criticality. *arXiv preprint arXiv:2508.11514* **2025**.
84. Ding, W.; Xu, C.; Arief, M.; Lin, H.; Li, B.; Zhao, D. A survey on safety-critical driving scenario generation—a methodological perspective. *IEEE Transactions on Intelligent Transportation Systems* **2023**, *24*, 6971–6988.
85. Zhou, R.; Huang, H.; Zhang, G.; Zhou, H.; Bian, J. Crash-Based Safety Testing of Autonomous Vehicles: Insights From Generating Safety-Critical Scenarios Based on In-Depth Crash Data. *IEEE Transactions on Intelligent Transportation Systems* **2025**.
86. Zhou, R.; Lin, Z.; Huang, X.; Peng, J.; Huang, H. Testing scenarios construction for connected and automated vehicles based on dynamic trajectory clustering method. In Proceedings of the 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2022, pp. 3304–3308.
87. Li, S.; Zhou, R.; Huang, H. Multidimensional Evaluation of Autonomous Driving Test Scenarios Based on AHP-EWN-TOPSIS Models. *Automotive Innovation* **2025**, pp. 1–15.
88. Wei, Z.; Huang, H.; Zhang, G.; Zhou, R.; Luo, X.; Li, S.; Zhou, H. Interactive critical scenario generation for autonomous vehicles testing based on in-depth crash data using reinforcement learning. *IEEE Transactions on Intelligent Vehicles* **2024**.
89. Luo, X.; Wei, Z.; Zhang, G.; Huang, H.; Zhou, R. High-risk powered two-wheelers scenarios generation for autonomous vehicle testing using WGAN. *Traffic Injury Prevention* **2025**, *26*, 243–251.
90. Wei, Z.; Zhou, H.; Zhou, R. Risk and Complexity Assessment of Autonomous Vehicle Testing Scenarios. *Applied Sciences* **2024**, *14*, 9866.
91. Wei, Z.; Bian, J.; Huang, H.; Zhou, R.; Zhou, H. Generating risky and realistic scenarios for autonomous vehicle tests involving powered two-wheelers: A novel reinforcement learning framework. *Accident Analysis & Prevention* **2025**, *218*, 108038.